ランダムなビット列における連 (証明論と論理・計算の構造)

AUTHOR(S):

川村, 保敬; 鈴木, 登志雄

CITATION:

川村, 保敬 ...[et al]. ランダムなビット列における連 (証明論と論理・計算の構造). 数理解析研究所講究録 2009, 1635: 23-32

ISSUE DATE:

2009-04

URL:

http://hdl.handle.net/2433/140465

数理解析研究所講究録
第1635巻 2009年 23-32

ランダムなビット列における連

川村保敬 (Yasutaka Kawamura)[1)]・鈴木 登志雄 (Toshio Suzuki)[2)]
首都大学東京 理工学研究科 数理情報科学専攻
(Departmet of Mathematics and Information Sciences,
Tokyo Metropolitan University)
1): kawamura-yasutaka@ed.tmu.ac.jp, 2): toshio-suzuki@tmu.ac.jp

平成20年12月5日

概 要

ビット列において，同じ文字が連続して現れる部分で極大なるものを連（run）という．与えられたオラクル X を無限ビット列とみなし，その最初の n ビットの始切片において，連の総数に対する長さ i の連の個数の比率に注目し，X と i を固定したまま n を無限大に飛ばして上記比率の極限を考察する．

(1) マーティンレフ・ランダムなオラクルにおいて，上記比率の極限が2の $-i$ 乗であることを観察する．

(2) ブール決定木（k ラウンドのAND-OR木）のコピーを並べた系列を葉から根への写像とみなす．この写像によってマーティンレフ・ランダムなオラクルから得られるオラクルにおいて，上記比率の極限が以下の通りとなることを示す．

$$\frac{p_k^{i-1}(1-p_k)+p_k(1-p_k)^{i-1}}{2},$$

ただし p_k は [Liu-Tanaka, *Inform. Process. Lett.* 2007] で与えられた確率であり，以下の漸化式で定まる．

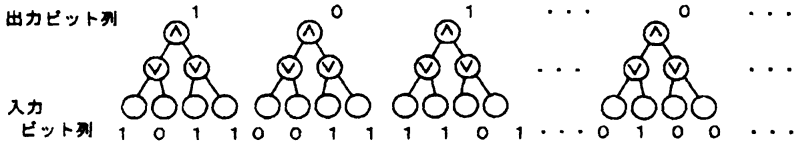
$$p_0=\frac{1}{2},\quad p_{k+1}=-p_k^4+2p_k^2.$$

(3) 擬似乱数を用いて行った実験結果にも触れる．

1 序

ブール決定木のリーフ（葉，leaf）に固定した真理値を与える代わりに真理値の確率分布を与えたものを「ランダム化されたブール決定木」という．この概念は [4] など多くの文献で研究されている．

鈴木は確率分布の代わりにマーティンレフ・ランダムなオラクルを与えた場合について考察した．すなわち，ブール決定木 可算無限個のコピーを，リーフのオラクルからルート（根，root）のオラクルへの写像と見なし，このような写像がランダム性を保存するか調べた．そして，マーティンレフのランダム性は保存されないが，その必要条件は保存されることを示した．[3]



我々は，鈴木のこの条件の他にも，どんなランダム性がこの種の写像で保存されるか，特に連についてどんな性質が保存されるかについて興味があり，研究を進めている．本稿で述べるのはこの研究の準備作業であり，ブール決定木による写像で写す前の，リーフ・ビット列における連の長さの分布と，写像した後のルート・ビット列における連の長さの分布についての報告である．

我々は，まず計算機実験を行って予想を立てた．疑似乱数によって長いビット列 $X(0), X(1), \cdots, X(n-1)$ を生成し，連の長さについて以下の近似式が成り立つことを観察した．ただし，i は n に比べて小さい自然数である．

$$\frac{X(0), X(1), \cdots, X(n-1) \text{における長さ} i \text{の連の個数}}{X(0), X(1), \cdots, X(n-1) \text{における連の個数}} \doteqdot \frac{1}{2^i}$$

そこで，「X がマーティンレフ・ランダムなオラクルで i が正の整数であるとき，上記の式左辺の極限（$n \to \infty$）が右辺に等しい」と予想した．この予想が成り立つことを第 4 節で示す．

また，k ラウンドの AND-OR 木によってマーティンレフ・ランダムなオラクルを写像して得られるオラクルにおいて，上記比率の極限が以下の通りとなることを示す．

$$\frac{p_k^{i-1}(1-p_k) + p_k(1-p_k)^{i-1}}{2}, \tag{1.1}$$

ただし p_k は k ラウンドの AND-OR 木の各リーフに，確率 $\frac{1}{2}$ ずつで値 1 と 0 をとる，一様で独立な確率分布を与えたときに，ルートが値 1 をとる確率である．この p_k の値は以下の漸化式で定まることが，[4] において示されている．

$$p_0 = \frac{1}{2}, \quad p_{k+1} = -p_k^4 + 2p_k^2.$$

第 4 節での議論を発展させることにより，第 5 節において (1.1) についての結果を示す．

第 2 節では用語と記号の説明を行い，第 3 節では実験について説明する．

2 用語と記号

非負整数全体の集合を ω で表す．長さ有限のビット列全体の集合を $\{0,1\}^*$ で表す．また，ω から $\{0,1\}$ への関数を**オラクル**といい，オラクル全体の集合を $\{0,1\}^\omega$ で表す．

ビット列111001111において，111と00および1111を連という．一般的な定義は次の通りである．

定義 1 ビット列において，同じ文字が連続して現れる部分で，極大となるものを，**連** *(run)* という．

次に，マーティンレフ・ランダム性について述べる．「ランダム性についての統計的検定のうち，計算機のプログラムで表せるようなものすべてに合格するようなオラクル」という概念の数学的モデルには，「計算機のプログラムで表せるようなもの」という部分をどう定式化するかに応じて様々な変種がある．その中でも代表的なものがマーティンレフ・ランダム性である [5].

定義 2 [2, Def.3.1] 集合族 $\mathcal{A} \subseteq \{0,1\}^\omega$ が**マーティンレフ零（*null*）集合**（あるいは Σ^0_1 **零集合**）であるとは，Σ^0_1 集合の一様に再帰的可算（recursively enumerable, あるいは computably enumerable）な列 $\{U_i\}_{i\in\omega}$ であって「$\forall i \in \omega(\mu(U_i) \leq 2^{-i})$」となるもの（**マーティンレフ・テスト**とよばれる）が存在して，「$\mathcal{A} \subseteq \bigcap_i U_i$」となることをいう．オラクル A について，$\{A\}$ が Σ^0_1 零集合でないとき，「A は**マーティンレフ・ランダム**（*Martin-Löf random*）である」，あるいは「*1* **ランダム**である」という．

上記定義における「Σ^0_1 集合の一様に再帰的可算な列 $\{U_i\}_{i\in\omega}$」という部分の意味は次の通りである．あるオラクル・チューリング機械 $M^{\sim}$（停止性についての保証はない）があって，任意のオラクル X と任意の自然数 i に対して以下が成り立つ．

$$X \in U_i \text{ if and only if } M^X(i) = 1.$$

マーティンレフ・ランダム性の定義は，以下に述べる構成的零集合（constructive null set）の概念を用いて特徴付けられる．

定義 3 [1, Def.6.26]

- 開集合 $G \subset \{0,1\}^\omega$ が**構成的開集合**（*constructively open set*）であるとは，ある再帰的可算な集合 $X \subset \{0,1\}^*$ に対して $G = X\{0,1\}^\omega$ となることをいう．
- 構成的開集合 $G_m = X_m\{0,1\}^\omega$ の列 $\{G_m\}_{m\geq 1}$ について以下の条件が成り立つとき，$\{G_m\}_{m\geq 1}$ を「**構成的開集合の構成的な列**(*constructive sequence of constructively open sets*, 略して *c.s.c.o. sets*)」という．「再帰的可算集合 $X \subset \{0,1\}^* \times \omega$ が存在して，すべての自然数 $m \geq 1$ に対して，

$$X_m = \{x \in \{0,1\}^* : (x,m) \in X\}$$

となる」.

- $S \subset \{0,1\}^\omega$ が**構成的零集合**（*constructively null set*）であるとは，c.s.c.o. sets $\{G_m\}_{m\geq 1}$ が存在して，以下が成り立つことをいう.

$$S \subset \bigcap_{m\geq 1} G_m \text{ かつ,}$$

$$\text{構成的に } \lim_{m\to\infty} \mu(G_m) = 0.$$

ただし，「構成的に $\lim_{m\to\infty} \mu(G_m) = 0$」とは，単調増加で有界でない計算可能な関数 $H : \omega \to \omega$ が存在して，任意の自然数 m, k に対し「$m \geq H(k)$ ならば $\mu(G_m) < 2^{-k}$」となることをいう.

定理 1 *[1]* X がマーティンレフ・ランダム $\Longleftrightarrow$ 任意の構成的零集合 S に対して，$X \notin S$ となる.

構成的零集合について，以下の結果が知られている．ここで「$\neq$」は「左辺が発散するか，または右辺と異なる値に収束する」ことを表す．これは，強い意味での大数の法則とみることができる.

定理 2 *[1, p.173, Theorem 6.27]* 以下の集合 Y は構成的零集合である.

$$Y = \{X \in \{0,1\}^\omega : \lim_{n\to\infty} \frac{X(0) + X(1) + \cdots + X(n-1)}{n} \neq \frac{1}{2}\}$$

定理 2 は，チェルノフ限界（Chernoff bound）を用いて証明される.

定理 3 チェルノフ限界 *[7, p.258, Lemma 11.9]* $X_0, X_1, \cdots, X_{n-1}$ の各々を，確率 p で 1 または 0 をとる独立な確率変数とする．このとき任意の θ（ただし $0 \leq \theta \leq 1$）に対して以下が成り立つ.

$$Prob[X_0 + X_1 + \cdots + X_{n-1} \geq (1+\theta)pn] \leq \exp\left(-\frac{\theta^2}{3}pn\right)$$

本論文では，構成的零集合の概念を利用して主要な結果を示す.

3 実験

ブール決定木のコピーを並べた系列をリーフのビット列からルートのビット列への写像とみなし，リーフとルートにおける連の長さの分布を調べる計算機実験を行った．実験は，以下の条件のもとに行われた.

- ブール決定木としては，図 1 にあるような AND-OR 木を用いた.

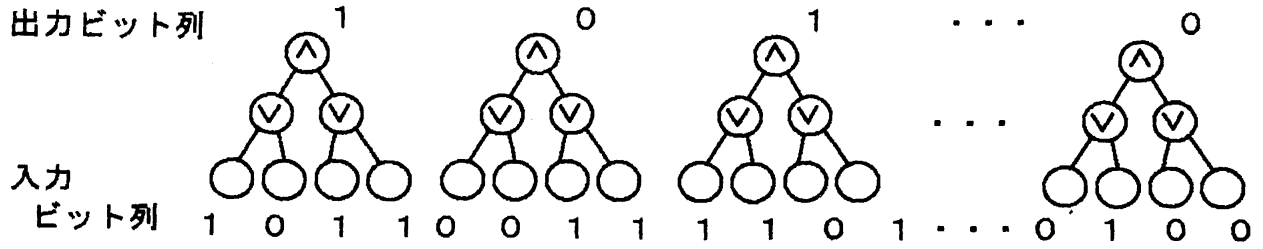


図 1: 木のコピーを並べた系列

- ルートからなるビット列が 128 ビットである場合について調べた.
- リーフには，疑似乱数を与えた.

木の深さ（depth）と似た概念としてラウンド（round）[4] というものを導入する. ブール決定木のコピーを並べた系列によってビット列を一回変換することを 1 ラウンドとよぶことにする. 図 2 は 2 ラウンドの木を並べた例である.

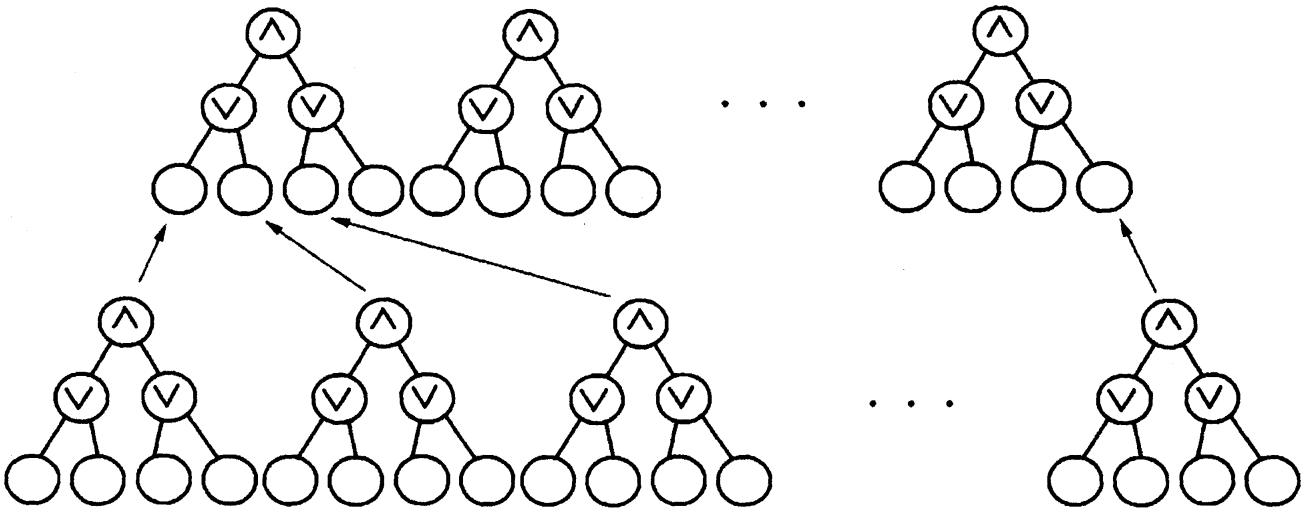


図 2: 2 ラウンドの木を並べた例

図 3，4 はそれぞれ，リーフ・ビット列とルート・ビット列（5 ラウンド）における連の長さの分布を表す. 時計の針の 3 時の位置から反時計回りの順に，(連の総数に対する）長さ 1 の連の（個数の）比率，長さ 2 の連の比率，$\cdots$ を表す.

疑似乱数によって生成されたリーフのビット列 $X(0), X(1), \cdots, X(n-1)$ において，以下の近似式が成り立つことが観察される. ただし，n は $2^{17}(= 4^5 \times 128)$ であり，i は n に比べて小さい自然数である.

$$\frac{X(0), X(1), \cdots, X(n-1) \text{における長さ} i \text{の連の個数}}{X(0), X(1), \cdots, X(n-1) \text{における連の個数}} \doteqdot \frac{1}{2^i}$$

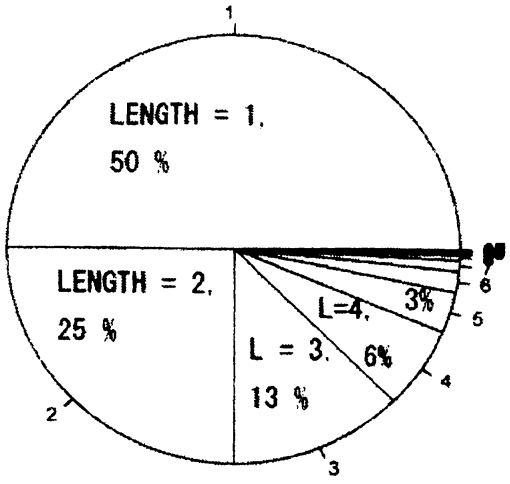


図 3: リーフの連の長さの分布

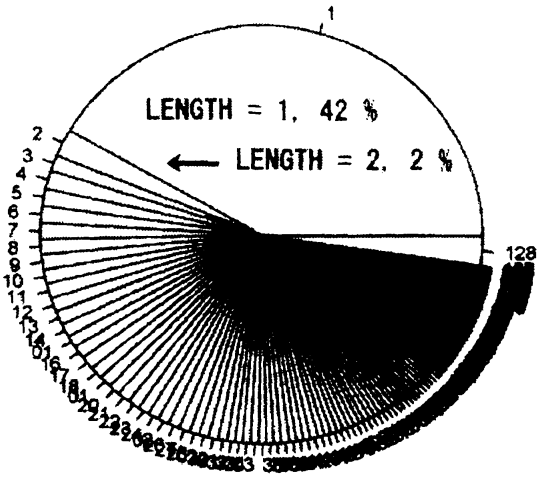


図 4: ルートの連の長さの分布（5 ラウンド）

4 マーティンレフ・ランダムなオラクルにおける連の分布

前節の実験に基いて，我々は以下の命題を予想した．

命題 4 X がマーティンレフ・ランダムなオラクルで i が自然数であるとき，以下が成り立つ．

$$\lim_{n\to\infty} \frac{X(0)X(1)\cdots X(n-1)\text{における長さ}\,i\,\text{の連の個数}}{X(0)X(1)\cdots X(n-1)\text{における連の個数}} = \frac{1}{2^i}$$

以下で上記を証明する．次の補題が証明の鍵となる．

補題 5 任意の正の整数 i に対し，以下の集合 $\mathcal{Y}$ は構成的零集合である．

$$\mathcal{Y} := \{X \in \{0,1\}^\omega : \lim_{n\to\infty} \frac{X(0)X(1)\cdots X(n-1)\text{における長さ}\,i\,\text{の連の個数}}{X(0)X(1)\cdots X(n-1)\text{における連の個数}} \neq \frac{1}{2^i}\}$$

上記補題を証明するため，準備をしよう．まず，与えられたオラクル X に対して，以下のように $r_n, r_{n,i}$ $(i = 1, 2, 3, \cdots)$ を定める．

$$\begin{aligned} r_n &:= (X(0)X(1)\cdots X(n-1)\text{における連の個数}), \\ r_{n,i} &:= (X(0)X(1)\cdots X(n-1)\text{における長さ}\,i\,\text{の連の個数}). \end{aligned}$$

さらにここで，

$$\begin{aligned} \mathcal{Y}_\infty &:= \{X \in \{0,1\}^\omega : \lim_{n\to\infty} \frac{r_n}{n} \neq \frac{1}{2}\}, \\ \mathcal{Y}_i &:= \{X \in \{0,1\}^\omega : \lim_{n\to\infty} \frac{r_{n,i}}{n} \neq \frac{1}{2^{i+1}}\}. \end{aligned}$$

とおく．これらがいずれも構成的零集合であることを示したい．

補題 6 $\mathcal{Y}_\infty$ は構成的零集合である.

証明 オラクル X と各々の自然数 j に対し，以下のように y_j を定める.

$$y_j = \begin{cases} 1 & \text{if } X(j) \text{が連の右端} \\ 0 & \text{otherwise} \end{cases}$$

ここで,「$y_j = 1$」となるための必要十分条件は「$X(j+1) \neq X(j)$」である．このとき，任意のオラクル X に対して以下が成り立つ．ただし，ここで「$\simeq$」は「左辺が収束するとき，かつ，そのときに限り右辺が収束して，そのとき両辺の値が一致する」ということを表す.

$$\lim_{n\to\infty} \frac{y_0 + y_1 + \cdots + y_{n-2}}{n-1} \simeq \lim_{n\to\infty} \frac{r_n}{n}$$

X の各ビットを独立に $\frac{1}{2}$ の確率で $0, 1$ のいずれかに決めるとき，j に関して独立に，y_j は $\frac{1}{2}$ の確率で $0, 1$ の値をとる．したがって，定理 2 の証明と同様にして（チェルノフ限界を用いて）$\mathcal{Y}_\infty$ が構成的零集合であることを示せる．Q.E.D.

補題 7 任意の構成的零集合 S_1，S_2 に対して，$S_1 \cup S_2$ も構成的零集合である.

証明 構成的零集合の定義にしたがって容易に確認できる．Q.E.D.

補題 8 任意の正の整数 i に対して，$\mathcal{Y}_i$ は構成的零集合である.

補題 8 もチェルノフ限界を用いて証明したいのであるが，チェルノフ限界は独立な事象についての定理である．そこで，無限ビット列としてのオラクルを長さ $i+2$ の区間に分割することによって，証明を独立な事象についての議論に還元する.

補題 8 の証明 まず $0 \leq s < i+2$ となる自然数 s を固定する.

$$s \leq j < (i+2)n + s \text{ かつ } j \equiv s \bmod (i+2) \tag{4.1}$$

となる自然数 j の各々に対して，$y_{s,j}$ を以下のように定める.

$$y_{s,j} = \begin{cases} 1 & \text{if } X(j+1) \text{が長さ } i \text{ の連の左端} \\ 0 & \text{otherwise} \end{cases}$$

ここで,「$y_{s,j} = 1$」となるための必要十分条件は

$$X(j) \neq X(j+1) = X(j+2) = \cdots = X(j+i) \neq X(j+i+1)$$

である．X の各ビットを独立に $\frac{1}{2}$ の確率で $0, 1$ のいずれかに決めるとき，j に関して独立に $y_{s,j}$ は確率 $\frac{1}{2^{i+1}}$ で値 1 をとる．よって，定理 2 と同様にチェルノフ限界を

用いて，以下の集合 $\mathcal{Y}_{i,s}$ が構成的零集合であることを示せる．ここで「$\neq$」は，左辺が発散するか，または右辺と異なる値に収束することを表す．

$$\mathcal{Y}_{i,s} := \{X : \lim_{n\to\infty} \frac{\sum_j y_{s,j}}{n} \neq \frac{1}{2^{i+1}}\}$$

ただし，総和記号は（s, n を固定して）(4.1) をみたすすべての j に渡る和である．以下の議論では，s の固定を解除する．

したがって補題 7 により，$\mathcal{Y}_{i,0} \cup \mathcal{Y}_{i,1} \cup \cdots \cup \mathcal{Y}_{i,i+1}$ は構成的零集合である．ここで，以下が成り立つ．ただし $\sharp$ は集合の濃度を表す．また，変数 k は非負整数を表し，「$\equiv$」は $i+2$ を法とした合同関係を表す．

$$(\mathcal{Y}_{i,0} \cup \mathcal{Y}_{i,1} \cup \cdots \cup \mathcal{Y}_{i,i+1})^c = \mathcal{Y}_{i,0}^c \cap \mathcal{Y}_{i,1}^c \cap \cdots \cap \mathcal{Y}_{i,i+1}^c$$

$$= \{X : 0 \leq \forall s < i+2 \quad \lim_{n\to\infty} \frac{\sum_j y_{s,j}}{n} = \frac{1}{2^{i+1}}\}$$

$$= \{X : 0 \leq \forall s < i+2$$

$$\lim_{n\to\infty} \frac{\sharp\{k < (i+2)n : X(k+1) \text{は長さ} i \text{の連の左端 かつ } k \equiv s\}}{n} = \frac{1}{2^{i+1}}\}$$

$$\subseteq \{X : \lim_{n\to\infty} \frac{\sharp\{k < (i+2)n : X(k+1) \text{は長さ} i \text{の連の左端}\}}{(i+2)n} = \frac{1}{2^{i+1}}\}$$

$$= \mathcal{Y}_i^c$$

したがって，$\mathcal{Y}_i$ は構成的零集合 $\mathcal{Y}_{i,0} \cup \mathcal{Y}_{i,1} \cup \cdots \cup \mathcal{Y}_{i,i+1}$ の部分集合であり，ゆえに $\mathcal{Y}_i$ は構成的零集合である．Q.E.D.

補題 5 の証明　$\mathcal{Y} \subset \mathcal{Y}_\infty \cup \mathcal{Y}_i$ であるから，補題 6 と補題 8 によって補題 5 が成り立つことがわかる．Q.E.D.

命題 4 の証明　補題 5 と定理 1 により，命題 4 が成り立つことがわかる．Q.E.D.

5　ML ランダムなオラクルを木で写像したオラクルにおける連の分布

定理 9　（主定理）　i と k は自然数であるとする．ブール決定木のコピーを無限個（ω 個）並べた系列を葉から根への写像とみなす．k ラウンドの *AND-OR* 木によってマーティンレフ・ランダムなオラクル X を写像して得られるルートのオラクル Y

に対して，以下が成り立つ．

$$\lim_{n\to\infty} \frac{Y(0)Y(1)\cdots Y(n-1)\text{における長さ}\,i\,\text{の連の個数}}{Y(0)Y(1)\cdots Y(n-1)\text{における連の個数}}$$
$$=\frac{p_k^{i-1}(1-p_k)+p_k(1-p_k)^{i-1}}{2},$$

ただし p_k は k ラウンドの *AND-OR* 木の各リーフに，確率 $\frac{1}{2}$ ずつで値 *1* と *0* をとる一様で独立な確率分布を与えたときに，ルートが値 *1* をとる確率である．この p_k の値は以下の漸化式で定まることが，*[4]* において示されている．

$$p_0=\frac{1}{2},\quad p_{k+1}=-p_k^4+2p_k^2.$$

以下で証明の概略を述べる．本節の残りの部分において，i,k は自然数とし，p は上記の p_k を表すものとする．また，与えられたオラクル X に対し，k ラウンドの AND-OR 木によって X を写像したオラクルを Y で表す．

まず，与えられたオラクル X と $a\in\{1,0\}$ に対して，以下のように $r_n^a, r_{n,i}^a$ $(i=1,2,3,\cdots)$ を定める．

r_n^a := $(Y(0)Y(1)\cdots Y(n-1)$ において，文字 a からなる連の個数$)$,
$r_{n,i}^a$:= $(Y(0)Y(1)\cdots Y(n-1)$ において，文字 a からなる長さ i の連の個数$)$.

また，オラクルのクラス $\mathcal{Z}, \mathcal{Z}^0, \mathcal{Z}^1$ を以下のように定める．

$$\mathcal{Z} := \{X\in\{0,1\}^\omega : \lim_{n\to\infty}\frac{r_{n,i}}{r_n}\neq\frac{p^{i-1}(1-p)+p(1-p)^{i-1}}{2}\},$$
$$\mathcal{Z}^0 := \{X\in\{0,1\}^\omega : \lim_{n\to\infty}\frac{r_{n,i}^0}{r_n^0}\neq p(1-p)^{i-1}\},$$
$$\mathcal{Z}^1 := \{X\in\{0,1\}^\omega : \lim_{n\to\infty}\frac{r_{n,i}^1}{r_n^1}\neq p^{i-1}(1-p)\}.$$

補題 10 任意のオラクル X に対して以下が成り立つ．
(1) $r_n=r_n^1+r_n^0$.
(2) $r_{n,i}=r_{n,i}^1+r_{n,i}^0$.
(3)

$$\lim_{n\to\infty} r_n=\infty \quad \text{ならば} \quad \lim_{n\to\infty}\frac{r_n^1}{r_n^0}=1.$$

証明 定義により，容易に確認できる．Q.E.D.

補題 11 *(1)* $\mathcal{Z}^1$ は構成的零集合である．
(2) $\mathcal{Z}^0$ は構成的零集合である．

証明の概略　(1) $\mathcal{Z}^1_\infty, \mathcal{Z}^1_i$ を以下のように定める.

$$\mathcal{Z}^1_\infty := \{X \in \{0,1\}^\omega : \lim_{n\to\infty} \frac{r^1_n}{n} \neq p(1-p)\},$$

$$\mathcal{Z}^1_i := \{X \in \{0,1\}^\omega : \lim_{n\to\infty} \frac{r^1_{n,i}}{n} \neq (1-p)^2 p^i\}$$

前節と同様の議論により，これら二つのクラスは構成的零集合であることから，$\mathcal{Z}^1$ が構成的零集合であることがわかる．つまり，(1) が成り立つ.

(2) の証明は (1) と同様である．Q.E.D.

定理 9（主定理）の証明　補題 11 により $\mathcal{Z}^0 \cup \mathcal{Z}^1$ は構成的零集合である．したがって定理 1 により，$\mathcal{Z}^0 \cup \mathcal{Z}^1$ の任意の要素はマーティンレフ・ランダムではない.

また，$\{r_n : n \in \omega\}$ が有界となる X 全体のクラスを $\mathcal{Z}^2$ とおく．$\mathcal{Z}^2$ の任意の要素は，再帰的（計算可能）であるから，マーティンレフ・ランダムではない.

ところが補題 10 により，$\mathcal{Z} \subseteq \mathcal{Z}^0 \cup \mathcal{Z}^1 \cup \mathcal{Z}^2$ が成り立つ．したがって定理 9 が成り立つ．Q.E.D.

なお，定理 9 は前節の実験結果（図 4）と整合する.

参考文献

[1] Calude, C.S.: *Information and randomness: an algorithmic perspective, 2nd ed.*. Springer, Berlin, 2002.

[2] Downey, R., Hirschfeldt, D. R., Nies, A. and Terwijn, S. A.: Calibrating randomness. *Bull. Symb. Log.*, **12**, pp.411-491 (2006).

[3] Kumabe, M. and Suzuki,T.: Weak randomness, genericity and Boolean decision trees., *Bull. Symb. Log.*, to appear. (An abstract of a talk at ALC10, the tenth Asian Logic Conference, 2008 September, Kobe.)

[4] Liu, C.-G. and Tanaka, K.: Eigen-distribution on random assignments for game trees. *Inform. Process. Lett.*, **104** pp.73-77 (2007)

[5] Martin-Löf, P:. The definition of random sequences. *Information and Control*, **9**, pp.602-619 (1966).

[6] Miller, J. S. and Nies, A.: Randomness and computability: open questions. *Bull. Symb. Log.*, **12**, pp.390-410 (2006).

[7] Papadimitriou, C. H.: *Computational complexity.* Addison-Wesley, Massachusetts (1994).